SUPPLEMENTARY INFORMATION

Genome-wide association study identifies new psoriasis susceptibility loci and an interaction between *HLA-C* and *ERAP1*

Amy Strange*, Francesca Capon*, Chris CA Spencer, Jo Knight, Michael E Weale, Michael H Allen, Anne Barton, Gavin Band, Céline Bellenguez, Judith GM Bergboer, Jenefer M Blackwell, Elvira Bramon, Suzannah J Bumpstead, Juan P Casas, Michael J Cork, Aiden Corvin, Panos Deloukas, Alexander Dilthey, Audrey Duncanson, Sarah Edkins, Xavier Estivill, Oliver Fitzgerald, Colin Freeman, Emiliano Giardina, Emma Gray, Angelika Hofer, Ulrike Hüffmeier, Sarah E Hunt, Alan D Irvine, Janusz Jankowski, Brian Kirby, Cordelia Langford, Jesús Lascorz, Joyce Leman, Stephen Leslie, Lotus Mallbris, Hugh S Markus, Christopher G Mathew, WH Irwin McLean, Ross McManus, Rotraut Mössner, Loukas Moutsianas, Åsa T Naluai, Frank O Nestle, Giuseppe Novelli, Alexandros Onoufriadis, Colin NA Palmer, Carlo Perricone, Matti Pirinen, Robert Plomin, Simon C Potter, Ramon M Pujol, Anna Rautanen, Eva Riveira-Munoz, Anthony W Ryan, Wolfgang Salmhofer, Lena Samuelsson, Stephen J Sawcer, Joost Schalkwijk, Catherine H Smith, Mona Ståhle, Zhan Su, Rachid Tazi-Ahnini, Heiko Traupe, Ananth C Viswanathan, Richard B Warren, Wolfgang Weger, Katarina Wolk, Nicholas Wood, Jane Worthington, Helen S Young, Patrick LJM Zeeuwen, Adrian Hayday, A David Burden, Christopher EM Griffiths, Juha Kere, André Reis, Gilean McVean, David M Evans, Matthew A Brown, Jonathan N Barker, Leena Peltonen, Peter Donnelly* and Richard C Trembath*

*These authors contributed equally to this work

Correspondence should be addressed to: peter.donnelly@well.ox.ac.uk or richard.trembath@kcl.ac.uk

SUPPLEMENTARY TABLES

Supplementary Table 1: Diagnostic criteria used in the collections

Discovery sample		
Dataset	*Case definition*	*N**
UK-London	Psoriasis diagnosed by a dermatologist[1]	873
UK-Manchester	Psoriasis diagnosed by a dermatologist[2]	725
IRE-Dublin	Psoriasis diagnosed by a dermatologist[3]	417
UK-Sheffield	Psoriasis diagnosed by a dermatologist, PsA patients excluded[4]	368
UK-Glasgow	Psoriasis diagnosed by a dermatologist[1]	352
	Control definition	*N*
1958 Birth Cohort	Population controls; unknown psoriasis status[5]	2,930
UK Blood Service	Population controls; unknown psoriasis status[5]	2,737

Replication sample		
Dataset	*Case definition*	*N*
GER-Erlangen	Psoriasis diagnosed by dermatologist, PsA patients excluded[6]	1,017
SWE-Stockholm	Psoriasis diagnosed by dermatologist, > 15 years of age[7]	706
AU-Graz	Psoriasis diagnosed by dermatologist, ≥ 18 years[8]	604
ITA-Rome	Psoriasis diagnosed by dermatologist, PsA patients excluded[9]	380
SPA-Barcelona	Psoriasis diagnosed by dermatologist[9]	357
SWE-Gothenburg	Psoriasis diagnosed by dermatologist[10]	262
HOL-Nijmegen	Psoriasis diagnosed by a dermatologist, PsA patients excluded[9]	237
UK-London	Psoriasis diagnosed by a dermatologist[1]	335
	Control definition	
UK-PoBI	Population controls; unknown psoriasis status[11]	2,866
GER-Erlangen	No history or family history of psoriasis[6]	930
HOL-Nijmegen	No history of psoriasis	463
SWE-Stockholm	No history of psoriasis[7]	454
AU-Graz	No history or family history of psoriasis, ≥ 18 years[8]	445
ITA-Rome	No history of psoriasis or other autoimmune disease[9]	380
SWE-Gothenburg	Population controls; unknown psoriasis status[10]	356
SPA-Barcelona	Population controls; unknown psoriasis status[9]	268

*N** refers to the number of cases contributed by each group, before DNA QC and *N* to cases and controls prior to genotyping QC. With the exception of four collections, case datasets included patients that were also affected by psoriatic arthritis (PsA), an inflammatory joint disease associated with the more severe forms of psoriasis.

Supplementary Table 2: Replication of the loci described in table 2 in each replication population

The numbers in brackets below the country show the number of cases and controls from that population. Each cell of the table shows the P value and below it the OR with the 95% confidence interval in brackets. [a] For rs702873, rs17716942 and rs465969 2,717 UK controls were used, and for rs7428395 there were no UK controls.

		1q36 rs4649203	2p16 rs702873	2q24 rs17716942	3p24 rs7428395	5q15 rs27524	6q21 rs240993	6q21 rs465969	14q13 rs8016947	19p13 rs12720356	19p13 rs280519
Austria (414, 328)	P	0.023	0.982	0.066	0.117	0.115	0.064	0.045	0.758	0.682	0.384
	OR	1.31 (1.04-1.66)	1.00 (0.82-1.23)	1.35 (0.98-1.85)	0.81 (0.63-1.05)	1.20 (0.96-1.52)	1.24 (0.99-1.56)	1.43 (1.01-2.04)	0.97 (0.79-1.19)	1.09 (0.71-1.67)	0.91 (0.73-1.13)
Sweden (945, 636)	P	0.257	0.425	0.044	0.919	0.799	0.051	0.250	$1.37x10^{-03}$	$7.36x10^{-04}$	0.095
	OR	1.10 (0.94-1.28)	0.94 (0.81-1.09)	1.24 (1.01-1.52)	1.01 (0.84-1.21)	1.02 (0.88-1.19)	1.17 (1.00-1.36)	1.16 (0.90-1.49)	1.26 (1.09-1.45)	1.52 (1.19-1.94)	1.13 (0.98-1.30)
Germany (988, 873)	P	0.532	$2.22x10^{-03}$	$6.75x10^{-03}$	$3.28x10^{-04}$	0.278	0.023	0.029	$1.01x10^{-03}$	0.107	0.060
	OR	1.05 (0.91-1.21)	1.23 (1.08-1.40)	1.32 (1.08-1.61)	1.38 (1.16-1.64)	1.08 (0.94-1.24)	1.18 (1.02-1.35)	1.27 (1.03-1.58)	1.25 (1.10-1.43)	1.25 (0.95-1.66)	1.14 (0.99-1.30)
Italy (351, 352)	P	0.305	0.788	0.044	0.086	0.045	$7.42x10^{-05}$	0.013	0.043	0.670	0.469
	OR	1.13 (0.90-1.42)	1.03 (0.83-1.27)	1.32 (1.01-1.72)	1.26 (0.97-1.65)	1.29 (1.01-1.64)	1.61 (1.27-2.04)	1.65 (1.11-2.44)	1.25 (1.01-1.54)	1.09 (0.72-1.66)	1.08 (0.87-1.34)
Netherlands (152, 445)	P	0.108	$1.17x10^{-04}$	0.071	0.069	$9.39x10^{-03}$	0.028	0.053	0.603	0.050	0.079
	OR	1.29 (0.95-1.76)	1.69 (1.30-2.21)	1.48 (0.97-2.26)	1.35 (0.98-1.85)	1.45 (1.10-1.93)	1.37 (1.03-1.81)	1.54 (0.99-2.39)	1.07 (0.82-1.40)	1.85 (1.00-3.43)	1.26 (0.97-1.63)
Spain (325, 254)	P	0.351	$2.76x10^{-03}$	0.125	0.422	0.439	0.055	0.254	0.021	0.347	0.019
	OR	1.14 (0.86-1.51)	1.44 (1.13-1.82)	1.29 (0.93-1.78)	1.13 (0.84-1.51)	1.11 (0.85-1.45)	1.30 (0.99-1.71)	1.29 (0.83-2.00)	1.32 (1.04-1.67)	1.27 (0.77-2.07)	1.35 (1.05-1.72)
UK (299, 2,578)[a]	P	0.113	0.866	0.136	-	0.089	0.028	$5.46x10^{-04}$	0.321	$2.63x10^{-04}$	0.129
	OR	1.17 (0.96-1.43)	1.01 (0.86-1.20)	1.22 (0.94-1.57)	-	1.16 (0.98-1.39)	1.23 (1.02-1.48)	1.63 (1.24-2.15)	1.09 (0.92-1.30)	2.02 (1.38-2.94)	1.14 (0.96-1.35)
Meta	P	$1.36x10^{-03}$	$1.41x10^{-03}$	$3.82x10^{-07}$	$4.92x10^{-03}$	$7.96x10^{-04}$	$3.37x10^{-09}$	$9.59x10^{-08}$	$7.89x10^{-07}$	$8.82x10^{-07}$	$5.93x10^{-04}$
	OR	1.13 (1.05-1.22)	1.12 (1.04-1.20)	1.29 (1.17-1.43)	1.14 (1.04-1.26)	1.13 (1.05-1.22)	1.25 (1.16-1.34)	1.37 (1.22-1.54)	1.19 (1.11-1.27)	1.40 (1.23-1.60)	1.13 (1.05-1.21)

Supplementary Table 3 Replication results for other SNPs taken into Sequenom genotyping

The three SNPs below the line are imputed (see Methods) and chosen from the imputed scan. Where the replication P value is '–', the replication risk is in the opposite direction from that in the discovery data. [a]rs7860566 was chosen as a proxy for rs9406469 r^2=0.87 calculated from 58C (r^2=0.93 from HapMap CEU)

Chr	rsID	Position	Risk allele	P_{scan}	P_{repl}
2p14	rs3845820	65829748	A	2.41×10^{-6}	-
5p13	rs896117	41035176	A	6.90×10^{-6}	-
8q13	rs9650151	66791406	A	4.69×10^{-6}	0.940
8q13	rs17396080	66817257	G	4.52×10^{-5}	-
9p23	rs7860566[a]	13843713	A	NA	0.104
9p23	rs9406469	13845745	A	6.13×10^{-6}	0.101
9q32	rs4574921	116578155	A	6.48×10^{-5}	8.53×10^{-3}
16p13	rs12232403	13274550	A	9.80×10^{-6}	0.380
19q13	rs8103587	38797218	G	3.16×10^{-5}	0.221
19q13	rs2059876	38822176	G	6.64×10^{-7}	0.316
19q13	rs11665818	44460056	G	2.72×10^{-4}	0.876
2p24	rs6736116	15282180	C	2.42×10^{-8}	0.280
19q13	rs516246	53897984	T	4.08×10^{-6}	0.069
19q13	rs492602	53898229	G	3.82×10^{-6}	0.142

Supplementary Table 4 Evidence for multiple signals of association at 2p16, 6q21 and 19p13

Odds ratios at each locus are estimated from a logistic regression model which includes the listed SNPs as explanatory variables. The P values relate to comparing the model with all the listed SNPs at the locus to a model without the SNP in question. At 19p13 only two of the SNPs were included in the replication genotyping. At 2p16 rs13014803 was not included in the replication genotyping.

		Discovery		Replication	
Chr	rsID	P values	OR (95%CI)	P values	OR (95%CI)
2p16	rs702873	3.93×10^{-5}	1.15 (1.06-1.26)	-	-
2p16	rs13014803	5.87×10^{-4}	1.21 (1.07-1.37)	-	-
6q21	rs240993	2.39×10^{-6}	1.25 (1.14-1.37)	2.08×10^{-4}	1.18 (1.08-1.28)
6q21	rs458017	4.23×10^{-4}	1.33 (1.13-1.55)	5.67×10^{-3}	1.21 (1.06-1.39)
19p13	rs280519	3.45×10^{-4}	1.15 (1.07-1.24)	0.0283	1.08 (1.01-1.16)
19p13	rs12720356	0.0363	1.17 (1.01-1.37)	1.38×10^{-5}	1.36 (1.18-1.56)
19p13	rs2278442	8.32×10^{-3}	1.12 (1.03-1.22)	-	-

Supplementary Table 5: Discovery sample interaction results

P values from tests (as described in text) of all pair-wise SNP interactions from the SNPs in Tables 1 and 2, where loci names are given and SNPs are ordered as in these tables. The *ERAP1 HLA-C* interaction P value is highlighted.

	LCE3D	IL13	TNIP1	IL12B	HLA-C	TNFAIP3	IL23A	ZNF313	IL28RA	REL	IFIH1	None rs6809854	ERAP1	TRAF3IP2	TRAF3IP2	NFKBIA	TYK2	TYK2
IL23R	0.32	0.58	0.93	0.95	0.50	0.61	0.26	0.88	0.39	0.81	0.91	0.93	0.41	0.67	0.75	0.63	0.45	0.26
LCE3D		0.15	0.23	0.87	0.08	0.38	0.83	0.88	0.61	0.83	0.47	0.60	0.29	0.15	0.40	0.94	0.23	0.12
IL13			0.12	0.76	0.16	0.51	0.70	0.65	0.82	0.14	0.28	0.47	0.75	0.67	0.90	0.81	0.35	0.31
TNIP1				0.82	0.34	0.89	0.70	0.66	0.52	0.27	0.86	0.55	0.47	0.73	0.98	0.98	0.87	0.70
IL12B					0.71	0.21	0.48	0.36	0.52	0.35	0.92	0.22	0.15	0.48	0.76	0.57	0.33	0.20
HLA-C						0.46	0.54	0.85	0.74	0.56	0.75	0.09	2.45×10^{-5}	0.32	0.35	0.77	0.93	0.76
TNFAIP3							0.50	0.67	0.78	0.60	0.03	0.86	0.39	0.61	0.76	0.69	0.26	0.50
IL23A								0.26	0.96	0.05	0.09	0.28	0.53	0.78	0.83	0.31	0.85	0.48
ZNF313									0.22	0.72	0.57	0.74	0.92	0.39	0.55	0.18	0.06	0.86
IL28RA										0.13	0.38	0.01	0.15	0.21	0.40	0.86	0.83	0.41
REL											0.69	0.33	0.69	0.80	0.48	0.90	0.60	0.82
IFIH1												0.66	0.01	0.28	0.84	0.73	0.86	0.32
None rs6809854													0.49	0.22	0.49	0.62	0.30	0.43
ERAP1														0.86	0.17	0.29	0.61	0.14
TRAF3IP2															0.60	0.27	0.24	0.49
TRAF3IP2																0.72	0.28	0.51
NFKBIA																	0.13	0.17
TYK2																		0.91

Supplementary Table 6: GWAS and replication results for samples stratified by *HLA-C* genotype

Chr	rsID	Position	Risk Allele	P_{scan}	P_{repl}	P_{comb}	Candidate gene
GWAS and replication results for individuals carrying the *HLA-C* risk allele at rs10484554							
2q11	rs17695937	97694307	A	$3.97x10^{-06}$	$7.13x10^{-03}$	$2.37x10^{-07}$	*ZAP70*
2q23	rs289858	151760572	G	$1.33x10^{-05}$	0.123	$3.20\ x10^{-05}$	-
4p15	rs17515558	12484449	A	$3.08x10^{-06}$	0.087	$7.52x10^{-06}$	-
5q15	rs27524	96127700	A	$3.00x10^{-09}$	$7.84x10^{-04}$	$4.24x10^{-11}$	*ERAP1*
13q22	rs9600888	76906155	G	$4.31x10^{-06}$	0.049	$3.29x10^{-06}$	*SCEL*
GWAS and replication results for individuals not carrying the *HLA-C* risk allele at rs10484554							
3p12	rs1437055	86913767	A	$2.16x10^{-05}$	0. 103		
3p12	rs6779214	86914729	G	$3.25x10^{-05}$	0.112		
7q36	rs6464341	152825093	G	$1.29x10^{-05}$	-		
10p13	rs11818063	15111624	G	$7.72x10^{-06}$	-		
12p13	rs11055463	7770921	A	$2.54x10^{-06}$	-		
15q26	rs1457853	93325178	A	$1.98x10^{-06}$	0.225		
16q22	rs7186310	65863014	C	$5.77x10^{-07}$	0.938		
16q22	rs11859352	65947949	G	$3.44x10^{-07}$	-		
17q11	rs11868086	25351708	G	$2.23x10^{-04}$	0.976		
17q11	rs12948898	25434450	A	$3.41x10^{-08}$	0.626		
18p11	rs8098483	13844080	G	$3.49x10^{-06}$	0.775		

Where the replication P value is '–', the replication risk in opposite direction to the discovery data.

Supplementary Table 7: Discovery sample interaction results using different SNPs

Interaction model comparison with three parameters where a dominant model is fitted at the HLA (β1), an additive model at the *ERAP1* SNP (β2) and an interaction term (β3), as described in methods. Lower residual deviance indicates improved model fit.

HLA-C	*ERAP1* SNP	Estimates			P values for β=0			Residual Deviance
		β1	β2	β3	β1	β2	β3	
rs10484554	rs27524	1.475	0.016	0.350	6.11×10^{-62}	0.790	5.64×10^{-05}	6995.2
rs10484554	rs30187	1.497	0.008	0.346	1.32×10^{-67}	0.897	9.08×10^{-05}	6997.6
Cw0602	rs27524	1.648	0.049	0.298	6.89×10^{-72}	0.388	8.28×10^{-04}	6896.0
Cw0602	rs30187	1.662	0.042	0.301	1.63×10^{-77}	0.468	9.04×10^{-04}	6897.1

Supplementary Table 8: Quality control exclusions

A) Individual sample QC showing numbers of individuals excluded where column headings are: QC: call rate and heterozygosity; Ancestry: HapMap PCA population exclusions; Relatedness: >5%IBD; Gender: gender modeled from X chromosome intensity mismatched with supplier gender; Intensity: outlying channel intensity; Identity: <90% concordant with initial Sequenom genotypes; Total: the number of unique samples excluded; Remaining: the number of samples carried through to the analysis stage. An individual is counted in each criteria it fails, thus can be included more than once on this table. B) Breakdown of SNP sample quality control exclusions Reasons for exclusions of SNPs. Pre-QC : Total number of SNPs; MAF : Minor Allele Frequency: HWE : Hardy Weinberg Equilibrium; Remaining - total SNPs after QC steps. A SNP is counted in each criteria it fails, thus can be included more than once on this table.

A

	QC	Ancestry	Relatedness	Gender	Intensity	Identity	Total	Remaining
PS	133	112	203	16	3	11	444	2178
58C	163	57	19	11	32	1	256	2674
UKBS	111	51	52	14	23	8	236	2501

B

	Pre-QC	MAF	Info	HWE	Plate effect	Remaining
Chr 1-22	580030	19322	45839	12125	26605	523081
Chr X	14194	621	1768	-	-	12413

Supplementary Table 9: Other SNPs with GWAS p-value less than 10^{-4}

All SNPs which passed a genome wide significance threshold of 10^{-4} in the genome wide association scan, excluding the SNPs in known and replicated regions, as coloured red and green in Figure 1. The odds ratio (OR) is given for the allele shown.

Table available as a separate excel document.

References for tables

1. Quaranta, M. et al. Differential contribution of CDKAL1 variants to psoriasis, Crohn's disease and type II diabetes. *Genes Immun* **10**, 654-8 (2009).

2. Smith, R.L. et al. Polymorphisms in the IL-12beta and IL-23R genes are associated with psoriasis of early onset in a UK cohort. *J Invest Dermatol* **128**, 1325-7 (2008).

3. Zhao, Y. et al. Filaggrin null alleles are not associated with psoriasis. *J Invest Dermatol* **127**, 1878-82 (2007).

4. Vasilopoulos, Y. et al. Association analysis of the skin barrier gene cystatin A at the PSORS5 locus in psoriatic patients: evidence for interaction between PSORS1 and PSORS5. *Eur J Hum Genet* **16**, 1002-9 (2008).

5. The Wellcome Trust Case-Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-78 (2007).

6. Huffmeier, U. et al. Characterisation of psoriasis susceptibility locus 6 (PSORS6) in patients with early onset psoriasis and evidence for interaction with PSORS1. *J Med Genet* **46**, 736-44 (2009).

7. Wolk, K. et al. Excessive body weight and smoking associates with a high risk of onset of plaque psoriasis. *Acta Derm Venereol* **89**, 492-7 (2009).

8. Weger, W. et al. The angiotensin-converting enzyme insertion/deletion and the endothelin -134 3A/4A gene polymorphisms in patients with chronic plaque psoriasis. *Exp Dermatol* **16**, 993-8 (2007).

9. de Cid, R. et al. Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. *Nat Genet* **41**, 211-5 (2009).

10. Inerot, A. et al. Collecting a set of psoriasis family material through a patient organisation; clinical characterisation and presence of additional disorders. *BMC Dermatol* **5**, 10 (2005).

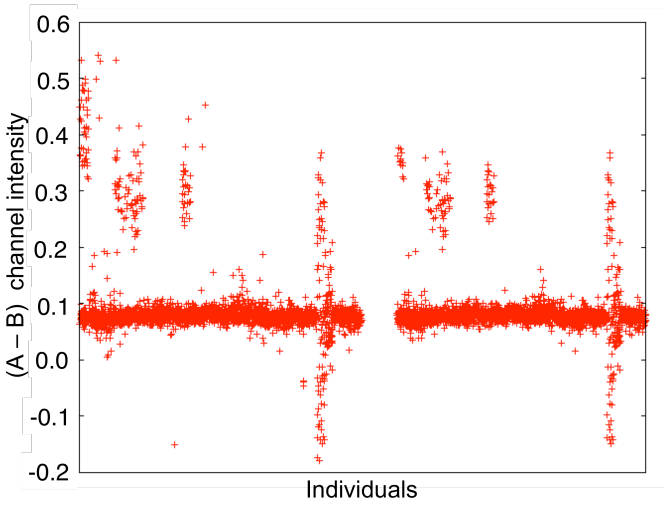
11. Barrett, J.C. et al. Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. *Nat Genet* **41**, 1330-4 (2009).

SUPPLEMENTARY FIGURES

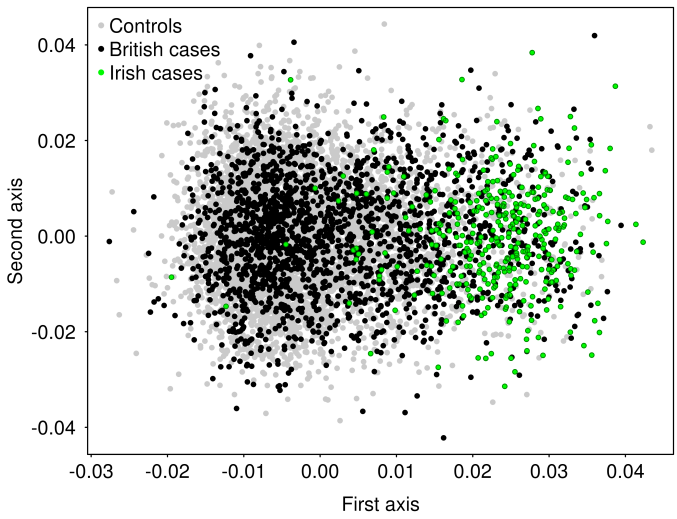
Supplementary Figure 1 Quality control

A) Individuals plotted on the mean difference between the A and B channel intensity. B) GWAS sample mapped on to the first two Principal Component Axes.The first two principal component scores based on a subset of 205,842 post-QC SNPs. The first principal component acts to differentiate British from Irish individuals and was used as a covariate in the logistic regression to control for population stratification. C) Quantile-quantile plot showing the chi squared distribution of autosomal SNPs after quality control (black), without the MHC (SNPs between 22 - 37 Mb removed - red) and without the loci implicated in psoriasis (blue), resulting in lambda of 1.045. The dotted lines show the 95% confidence interval for the quantiles of the chi squared distribution.

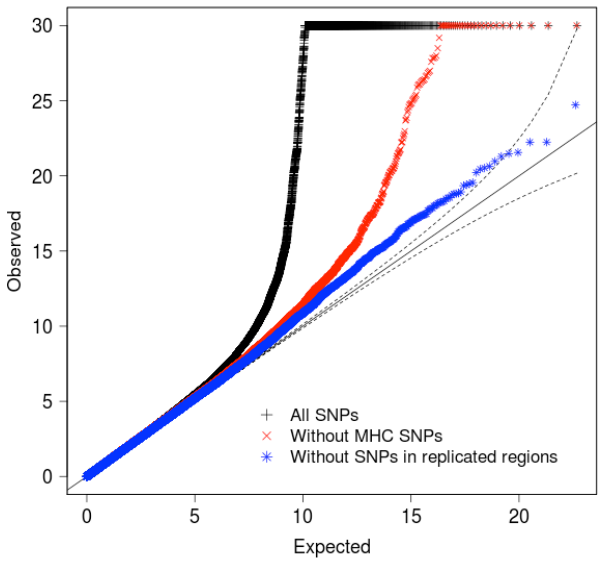
A



B



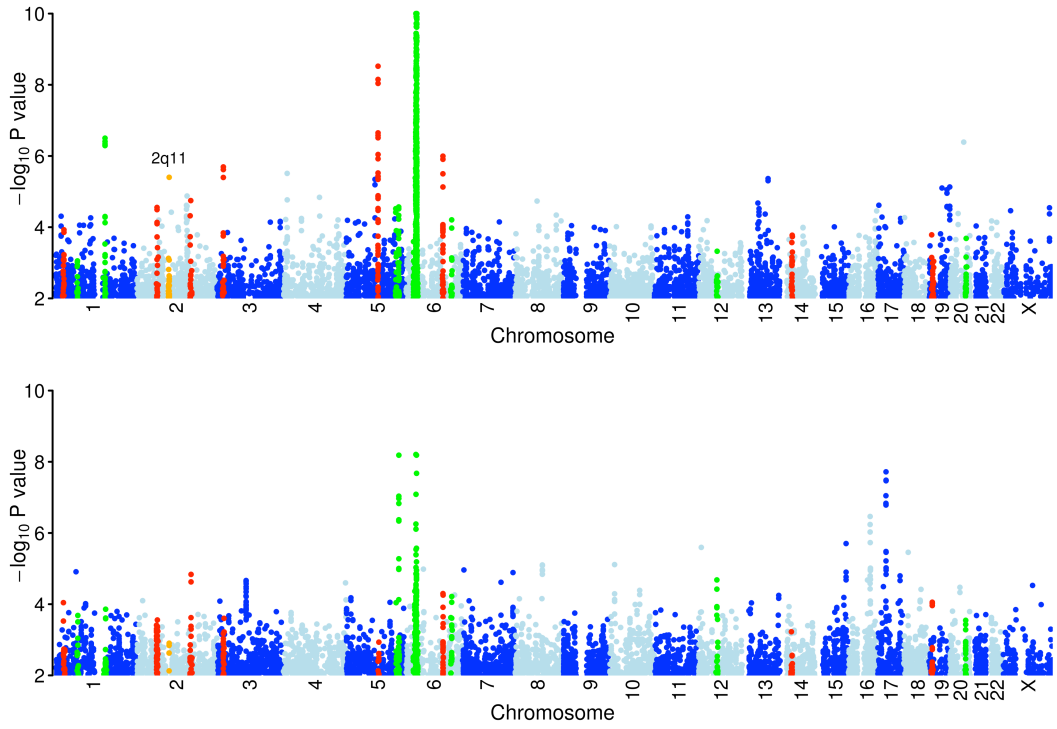
C



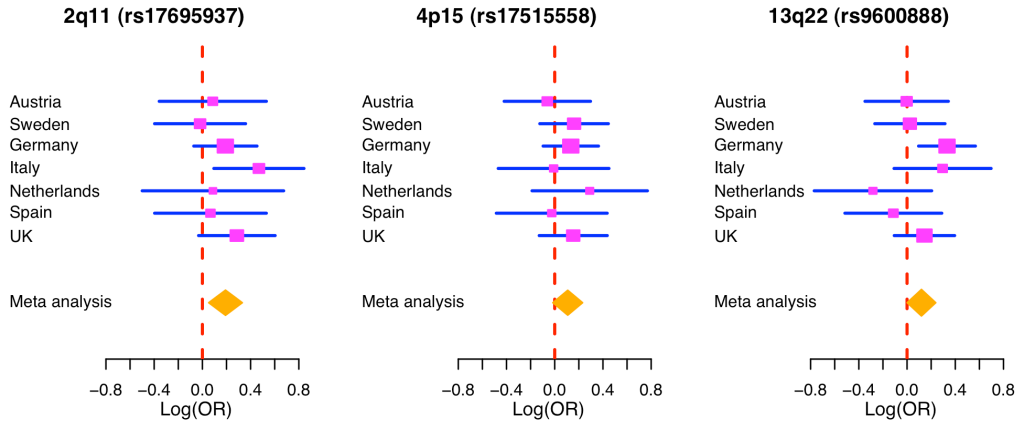
Supplementary Figure 2 Genome wide association results and replication forest plots for the analysis stratified by rs10484554

A) Genome wide association plots where individuals are stratified by *HLA-C* risk. Genome wide association results from 523,067 SNPs on chromosomes 1-22 and 12,408 SNPs on the X chromosome using the additive model in SNPtest. The $-\log_{10}$ (P) values are calculated from the 1-d.f. trend test and values are thresholded at 10^{-10}. The top plot shows results from an analysis of individuals who carry the risk allele at rs10484554. The bottom plot shows results from the analysis on individuals who do not carry the risk allele at rs10484554. Regions in red are described in table 2. Regions which have been shown previously to be associated with psoriasis and which replicate in this study are highlighted in green, as described in table 1. Regions which show strong suggestive evidence in the combined data are highlighted in orange. B) Forest plots for *HLA-C* analysis replication SNPs. Evidence for association at the three loci reaching borderline significance in individuals carrying the *HLA-C* risk allele at rs10484554 in the replication population. The blue lines show the 95% confidence intervals of the log(odds ratio) for each population. The diamond indicates the 95% confidence interval for the meta-analysis of these populations.

A

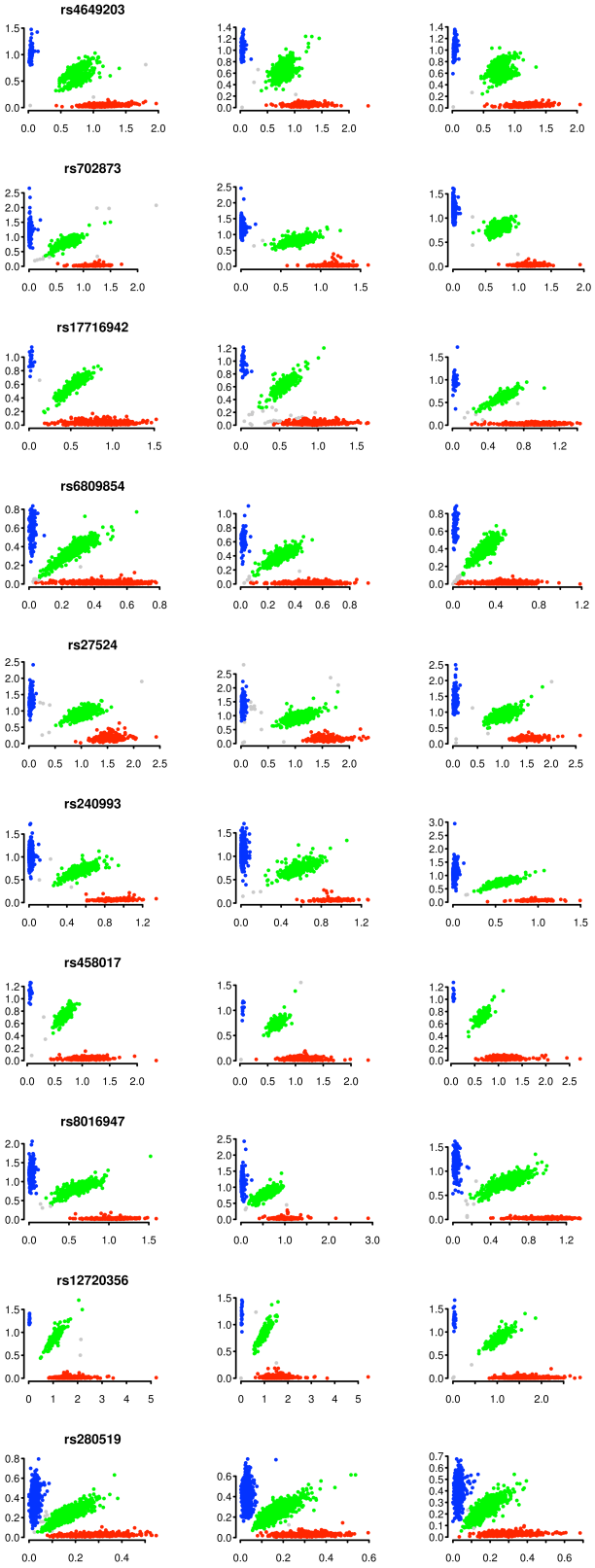


B



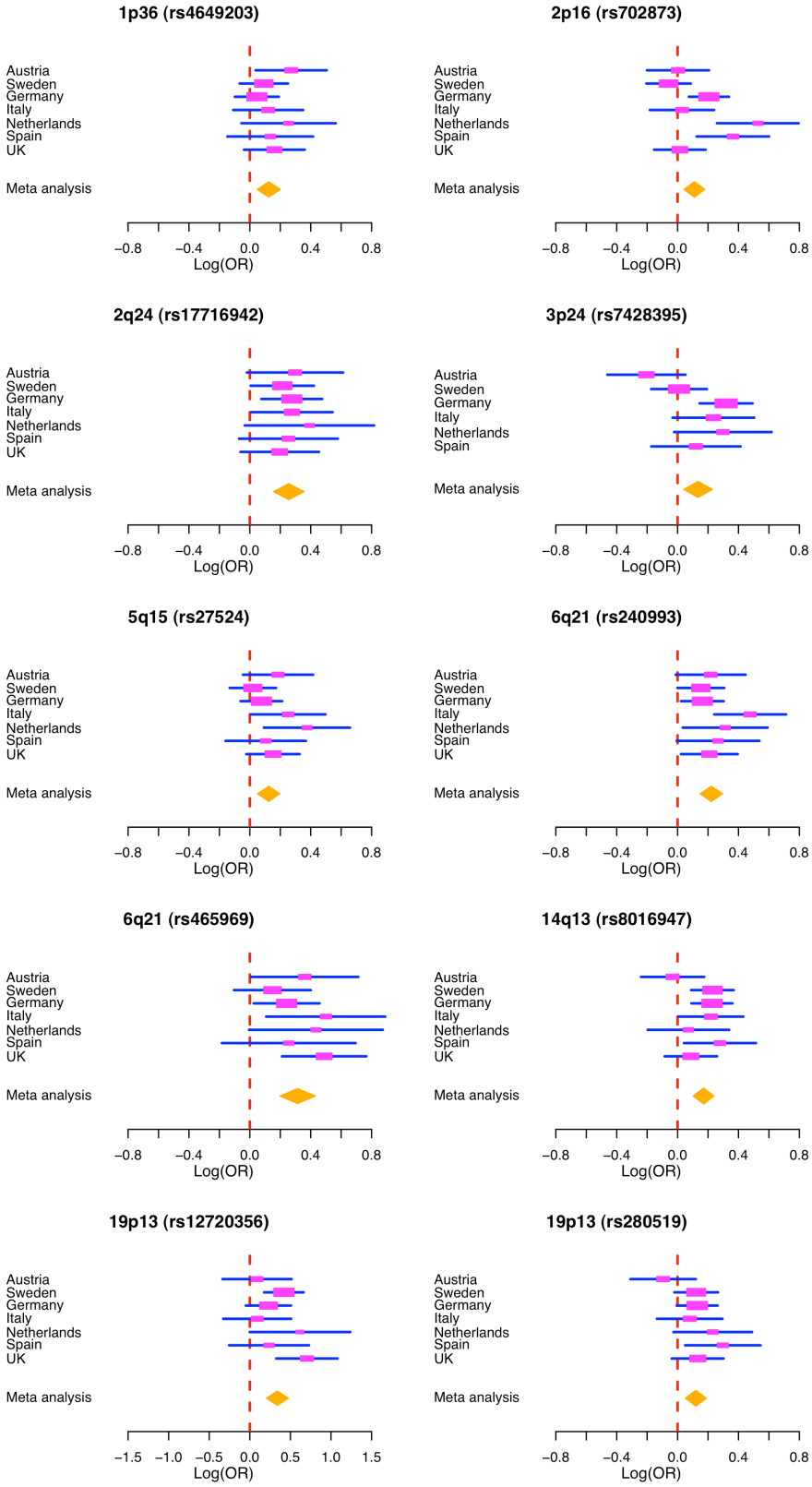
Supplementary Figure 3 Cluster plots of SNPs detailed in Table 2

The A and B channel intensities are plotted and coloured by genotype, where missing genotypes are grey, red and blue are AA and BB homozygotes and green are heterozygotes. There are three plots for each SNP showing from left to right, the genotypes of case individuals, 58C and UKBS.



Supplementary Figure 4 Forest plots for replication analysis

Evidence for association at the eight loci in the replication populations where the blue lines show the 95% confidence intervals of the log(odds ratio) for each population. The diamond indicates the 95% confidence interval for the meta-analysis of these populations.



SUPPLEMENTARY NOTE 1: Additional results and methodology

HLA analysis. As is typical, our study typed SNP variants. We undertook additional analyses to understand the relationship of the findings described above with the known classical HLA risk allele, Cw*0602. Based on molecular typing in a subset of 725 of the 58C individuals, the correlation between Cw*0602 genotypes (0, 1, or 2 copies of Cw*0602) and our top SNP, rs10484554, is r2= 0.7. Next, we undertook imputation of the *HLA-C* genotypes from our SNP data (see Methods). The Cw*0602 allele is imputed with high accuracy (false positive rate 1%, false negative rate < 1%). When looking at the overlap set of individuals for which the HLA-Cw*0602 imputation passed a confidence threshold of 0.7 and which have genotype calls at rs10484554, association testing with these imputed genotypes showed a stronger signal for the classical allele than for any single SNP. For the HLA Cw*0602 type, the P value is 5.44×10^{-221}, OR =5.55, 95%CI: 4.95-6.21, and for the top SNP the P value is 3.05×10^{-202}, OR =4.64, 95%CI: 4.17-5.15, but conditional analyses are not definitive in ruling out either variant as the primary source of the signal. The effect at Cw*0602 also approximates a dominant model (slightly more closely than does the top SNP: the risk for the heterozygote HLA type is 6.47, and for individuals homozygous for Cw*0602, the odds ratio is 9.91).

DNA sample preparation. Genomic DNA for all cases was shipped to the Sanger Institute, Cambridge. Where there was sufficient DNA, quality was validated using the Sequenom iPLEX assay designed to genotype four gender SNPs and 26 SNPs present on the Illumina Beadchips. DNA concentrations were quantified using a PicoGreen assay (Invitrogen) and an aliquot assayed by agarose gel electrophoresis. A DNA sample was considered to pass quality control if the DNA concentration was greater than or equal to 50 ng/μl, the DNA was not degraded, the gender assignment from the iPLEX assay matched that provided in the patient data manifest and genotypes were obtained for at least two thirds of the SNPs on the iPLEX.

Discovery data genotyping. Samples were genotyped at the Sanger Institute on the Illumina Infinium platform. Samples from the case collection were genotyped on the Human660W-Quad (a custom chip designed by WTCCC2 and comprising Human550 and a set of circa 6000 common CNVs from the Structural Variation Consortium[1]) and samples from the control collections were genotyped on the custom Human1.2M-Duo (a WTCCC2 custom array comprising Human1M-Duo and the common CNV content described above). Bead intensity data was processed and normalized

for each sample in BeadStudio; data for successfully genotyped samples was extracted and genotypes called within collections using Illuminus[2].

Preliminary analysis of the genotype data indicated a higher than expected number of SNPs showing sporadic signals of association (i.e. not compatible with local patterns of linkage disequilibrium). Inspection of cluster plots of these SNPs suggested a small number of individuals with noticeable but inconsistent shift in signal intensity which often led to incorrect genotype calling (Supplementary Figure 5). To investigate this we looked for genome-wide differences in the two channels (A and B) used to measure the presence of each allele at a SNP. As there is no population genetic reason for individuals to carry more A alleles rather than the B alleles, outlying individuals are likely to be due to genotyping artefacts. Through this analysis we identified 271 case samples with clear differences in the relative intensity of the two channels. Further investigation implicated both poorly performing chips and reagents (which were replaced by Illumina), as well as variability in lab protocols, as responsible for the outliers. We therefore believe our experience may usefully highlight the value of such checks for similar experiments, especially since the effects of these artefacts can be subtle, and difficult to detect by standard QC metrics. Accordingly, we re-genotyped these samples, together with all samples in plates where at least 30% of samples were intensity outliers (394 total samples). In addition, we introduced an additional QC step which removed any remaining intensity outliers from association analyses (see below).

Replication data genotyping. Genotyping was carried out at the Sanger Institute using the Sequenom iPLEX Gold assay. Individual samples were excluded from analysis if they had call rates <80% or if the reported gender was discordant with gender specific markers. We removed pairs of samples showing genotype concordance indicative of being duplicates. The PoBI samples were genotyped on the custom Human1.2M-Duo array using Illumina's Infinium platform. These data were subject to similar quality control criteria as described below and for each SNP used in the replication study the cluster plot was visually inspected.

Quality Control. *Samples:* Rather than attempting to model all the complexities of the sample collection and data generating process we identified and removed samples whose genome-wide patterns of diversity differed from the collection at large, interpreting them as likely to be due to biases or artefacts. To do so we used a Bayesian clustering approach to infer outlying individuals

on the basis of call rate, heterozygosity, relatedness and ancestry as previously described[3]. Additionally we applied the same clustering approach to raw intensity data to infer remaining signal intensity outliers. For each sample, the difference between the A channel intensity and the B channel intensity was averaged over 10,000 SNPs on chromosome 22 (chosen arbitrarily) and this data was modeled as a mixture of "normal" and "outlier" classes, with the outliers being excluded as above. The same approach was used taking intensity measures from the A channel on the non-pseudo autosomal X chromosomes to similarly identify outliers and infer gender. To guard against some possible sample mishandling, we removed samples if their inferred gender was discordant with the recorded gender, or if less than 90% of the SNPs typed by Sequenom on entry to sample handing (see above) agreed with the genome-wide data.

To obtain a set of putatively unrelated individuals we used a hidden Markov model (HMM) to infer identify by descent along the genome between pairs of individuals within each collection. Amongst pairs of closely related individuals, we excluded the individual with the lowest call rate, iteratively repeating this procedure to obtain a set of individuals with pairwise identity by descent less than 5%[3].

The above set of QC procedures resulted in 444 case individuals and 492 control individuals being excluded from the discovery dataset (see Supplementary Table 8 for further details). Of the 2,178 case samples which passed quality control, 345 have self-reported Irish ancestry.

SNPs: A measure of (Fisher) information for the allele frequency at each SNP was calculated using SNPTEST. Autosomal SNPs were excluded if in the case data or either of the control collections, this information measure was below 0.98, or if the minor allele frequency was less than 0.01%, or if the Hardy Weinberg P value was lower than 10^{-20} (unless the SNP was within the MHC region). Association between the SNP and the plate on which samples were genotyped was calculated and SNPs with a plate effect P value less than 10^{-6} were also excluded. The above filters removed 9.8% of the SNPs and a further 14 SNPs were removed after visual inspection of cluster plots, leaving 523,067 out of 580,030 autosomal SNPs for further analysis (Supplementary Table 8). Cluster plots for SNPs in Table 2 are shown in Supplementary Figure 3. QC for SNPs on the X chromosome was performed as just described, with the exception of the Hardy-Weinberg and plate-effect filters,

resulting in the removal of 12.5% of SNPs, with a further five SNPs removed upon inspection of cluster plots leaving 12,408 SNPs (Supplementary Table 8). The proportion of SNPs excluded is higher on the X chromosome because the genotyping algorithm is not as well calibrated to the pattern of heterozygosity.

Imputation: For the haploid reference panel we used HapMap2 and HapMap3 SNP data (www.hapmap.org) on the 120 non-related CEU trios, and for the diploid reference we used a merged set of genotype calls from Affymetrix 6.0 and Illumina custom Human1.2M-Duo genotyping chip typed on the 58C and UKBS individuals forming the common control group of WTCCC2. Prior to imputation the SNPs were subjected to additional, more stringent, QC to remove any SNPs with poorly called genotypes which could adversely affect the quality of the imputation.

Secondary signals: We used two methods to investigate possible secondary signals within GWAS association regions. The first was to perform standard frequentist conditional analyses. The second used the program GENECLUSTER[4] which adopts a Bayesian approach to look for primary and secondary association signals at known and putative SNPs. Where evidence for secondary signals existed we fitted logistic regression models including each implicated SNP as a predictor (as well as the principal component covariate). We used likelihood ratio tests to compare the models with and without the secondary SNP. The effect of carry an additional copy of each of the risk alleles was obtained by summing the estimated coefficients associated with each SNP and by using the estimated variance-covariance matrix to calculate confidence intervals on the log odds scale.

References for text

1. Conrad, D.F. et al. Origins and functional impact of copy number variation in the human genome. *Nature* (2009).
2. Teo, Y.Y. et al. A genotype calling algorithm for the Illumina BeadArray platform. Bioinformatics 23, 2741-6 (2007).
3. Barrett, J.C. et al. Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. Nat Genet 41, 1330-4 (2009).
4. Su, Z., Cardin, N., The Wellcome Trust Case Control Consortium, Donnelly, P. & Marchini, J. A Bayesian Method for Detecting and Characterizing Allelic Heterogeneity and Boosting Signals in Genome-Wide Association Studies. *Statistical Science* **24**, 430-450 (2009).

SUPPLEMENTARY NOTE 2: Consortia membership

Membership of Wellcome Trust Case Control Consortium 2

Management Committee

Peter Donnelly (Chair)[1,2], Leena Peltonen (Deputy Chair)[3], Jenefer M Blackwell[4, 5], Elvira Bramon[6] , Matthew A Brown[7] , Juan P Casas[8] , Aiden Corvin[9], Nicholas Craddock[10], Panos Deloukas[3], Audrey Duncanson[11], Janusz Jankowski[12], Hugh S Markus[13], Christopher G Mathew[14], Mark I McCarthy[15], Colin NA Palmer[16], Robert Plomin[17], Anna Rautanen[1], Stephen J Sawcer[18], Nilesh Samani[19], Richard C Trembath[14], Ananth C Viswanathan[20,21], Nicholas W Wood[22]

Data and Analysis Group

Chris C A Spencer[1], Gavin Band[1], Céline Bellenguez[1], Colin Freeman[1], Garrett Hellenthal[1], Eleni Giannoulatou[1], Matti Pirinen[1], Richard Pearson[1], Amy Strange[1], Zhan Su[1], Damjan Vukcevic[1], Peter Donnelly[1,2]

DNA, Genotyping, Data QC and Informatics Group

Cordelia Langford[3], Sarah E Hunt[3], Sarah Edkins[3], Rhian Gwilliam[3], Hannah Blackburn[3], Suzannah J Bumpstead[3], Serge Dronov[3], Matthew Gillman[3], Emma Gray[3], Naomi Hammond[3], Alagurevathi Jayakumar[3], Owen T McCann[3], Jennifer Liddle[3], Marc L Perez[3], Simon C Potter[3], Radhi Ravindrarajah[3], Michelle Ricketts[3], Matthew Waller[3], Paul Weston[3], Sara Widaa[3], Pamela Whittaker[3], Panos Deloukas[3], Leena Peltonen[3]

Publications Committee

Christopher G Mathew (Chair)[14], Jenefer M Blackwell[4,5], Matthew A Brown[7], Aiden Corvin[9], Mark I McCarthy[15], Chris C A Spencer[1]

1 Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7LJ, UK; 2 Dept Statistics, University of Oxford, Oxford OX1 3TG, UK; 3 Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; 4 Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, 100 Roberts Road, Suciaco, Western Australia 6008; 5 Genetics and Infection Laboratory, Cambridge Institute of Medical Research, Addenbrooke's Hospital, Cambridge CB2 0XY, UK; 6 Division of Psychological Medicine and Psychiatry, Biomedical Research Centre for Mental Health at the Institute of Psychiatry, King's College London and The South London and Maudsley NHS Foundation Trust, Denmark Hill, London SE5 8AF, UK; 7 Diamantina Institute of Cancer, Immunology and Metabolic Medicine, Princess Alexandra Hospital, University of Queensland, Brisbane, Queensland, Australia; 8 Dept Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK; 9 Neuropsychiatric Genetics Research Group, Institute of Molecular Medicine, Trinity College Dublin, Dublin 2, Eire; 10 Dept Psychological Medicine, Cardiff University School of Medicine, Heath Park, Cardiff CF14 4XN, UK; 11 Molecular and Physiological Sciences, The Wellcome Trust, London NW1 2BE; 12 Centre for Gastroenterology, Bart's and the London School of Medicine and Dentistry, London E1 2AT, UK; 13 Clinical Neurosciences, St George's University of London, London SW17 0RE; 14 Dept Medical and Molecular Genetics, King's College London School of Medicine, Guy's Hospital, London SE1

9RT, UK; 15 Oxford Centre for Diabetes, Endocrinology and Metabolism (ICDEM), Churchill Hospital, Oxford OX3 7LJ, UK; 16 Biomedical Research Centre, Ninewells Hospital and Medical School, Dundee DD1 9SY, UK; 17 Social, Genetic and Developmental Psychiatry Centre, King's College London Institute of Psychiatry, Denmark Hill, London SE5 8AF, UK; 18 University of Cambridge Dept Clinical Neurosciences, Addenbrooke's Hospital, Cambridge CB2 2QQ, UK; 19 Dept Cardiovascular Science, University of Leicester, Glenfield Hospital, Leicester LE3 9QP; 20 Glaucoma Research Unit, Moorfields Eye Hospital NHS Foundation Trust, London EC1V 2PD,UK; 21 Dept Genetics, University College London Institute of Ophthalmology, London EC1V 9EL, UK; 22 Dept Molecular Neuroscience, Institute of Neurology, Queen Square, London WC1N 3BG, UK.

Membership of GAP Consortium

Management Committee

Richard C Trembath (Chair)[1,2], Jonathan N Barker[1], A David Burden[3], Michael J Cork[4], Xavier Estivill[5], Christopher EM Griffiths[6], Juha Kere[7], Ross McManus[8], Giuseppe Novelli[9], André Reis[10], Lena Samuelsson[11], Joost Schalkwijk[12], Mona Ståhle[13], Rachid Tazi-Ahnini[4], Wolfgang Weger[14], Jane Worthington[15]

KCL group

Michael H Allen[1], Jonathan N Barker[1], Francesca Capon[1], Adrian Hayday[16], Jo Knight[1,2], Frank O Nestle[1], Alexandros Onoufriadis[1], Catherine H Smith[17], Richard C Trembath[1,2], Michael E Weale[1]

AU-Graz group:

Angelika Hofer[14], Wolfgang Salmhofer[14], Wolfgang Weger[14], Peter Wolf[14]

FIN-Helsinki group:

Kati Kainu[18], Juha Kere[7], Ulpu Saarialho-Kere[18], Sari Suomela[18]

GER-Erlangen group:

Petra Badorf[10], Ulrike Hüffmeier[10], Werner Kurrat[19], Wolfgang Küster[20], Jesús Lascorz[21], Rotraut Mössner[22], André Reis[10], Funda Schürmeier-Horst[23], Markward Ständer[24], Heiko Traupe[23]

HOL-Nijmegen group:

Judith G M Bergboer[12], Martin den Heijer[25], Joost Schalkwijk[12], Peter C. van de Kerkhof[12], Patrick L J M Zeeuwen[12]

IRE-Dublin group:

Louise Barnes[8], Linda E Campbell[26], Catriona Cusack[27], Oliver Fitzgerald[28], Phil Gallagher[28], Brian Kirby[28], Alan D Irvine[29], Trevor Markham[27], WH Irwin McLean[26], Ross McManus[8], Joe McPartlin[8], Sarah F Rogers[28], Anthony W Ryan[8], Agnieszka Zawirska[28]

ITA-Rome group:

Emiliano Giardina[9], Tiziana Lepre[9], Giuseppe Novelli[9], Carlo Perricone[9]

SPA-Barcelona group:

Xavier Estivill[5], Gemma Martín-Ezquerra[30], Ramon M Pujol[30], Eva Riveira-Munoz[5]

SWE-Gothenburg group:

Annica Inerot[31], Åsa T Naluai[11], Lena Samuelsson[11],

SWE-Stockholm group:

Lotus Mallbris[13], Mona Ståhle[13], Katarina Wolk[13]

UK-Glasgow group:

A David Burden[3], Joyce Leman[3]

UK-Manchester group:
Anne Barton[15], Christopher EM Griffiths[6], Richard B Warren[6], Jane Worthington[15], Helen S Young[6]
UK-Sheffield group:

Michael J Cork[4], Rachid Tazi-Ahnini[4]

[1]Division of Genetics and Molecular Medicine, King's College London, London, UK; [2]National Institute for Health Research (NIHR), Biomedical Research Centre, Guy's and St. Thomas' NHS Foundation Trust and King's College London, London, UK; [3]Department of Dermatology, Western Infirmary, Glasgow, UK; [4]Academic Unit of Dermatology Research, Department of Infection and Immunity, The University of Sheffield, Sheffield, UK; [5]Genes and Disease Programme, Centre for Genomic Regulation (CRG) and Public Health and Epidemiology Network Biomedical Research Center (CIBERESP), Barcelona, Spain; [6]Dermatological Sciences, Salford Royal NHS Foundation Trust, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK; [7]Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden, and Folkhälsan Institute of Genetics, Helsinki, Finland, and Department of Medical Genetics, University of Helsinki, Finland; [8]St. James's Hospital, Trinity College, Dublin, Ireland; [9]Department of Biopathology, Centre of Excellence for Genomic risk Assessment in Multifactorial and Complex Diseases, School of Medicine, University of Rome Tor Vergata, Italy; [10]Institute of Human Genetics, University of Erlangen-Nuremberg, Erlangen, Germany; [11]Department of Medical and Clinical Genetics, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; [12]Department of Dermatology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; [13]Dermatology Unit, Department of Medicine, Karolinska Institutet, Stockholm, Sweden; [14]Department of Dermatology, Medical University of Graz, Graz, Austria; [15]Arthritis Research UK Epidemiology Unit, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK; [16]Division of Immunology, Infection and Inflammatory Disease; King's College London, London, UK; [17]St John's Institute of Dermatology, King's College London, London, UK; [18]Department of Dermatology and Venerology, University of Helsinki, Helsinki, Finland; [19]Asklepios Nordseeklinik, Westerland/Sylt, Germany; [20]TOMESA Clinics, Bad Salschlirf, Germany; [21]Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany; [22]Department of Dermatology, University of Göttingen, Göttingen, Germany; [23]Department of Dermatology, University of Münster, Münster, Germany; [24]Psoriasis Rehabilitation Hospital, Bad Bentheim, Germany; [25]Department of Endocrinology and Department of Epidemiology and Biostatistics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; [26]University of Dundee, Dundee, UK; [27]University College Hospital Galway, Galway, Ireland; [28]St Vincent's University Hospital, Dublin, Ireland; [29]Department of Clinical Medicine, Trinity College Dublin, Our Lady's Children's Hospital Crumlin, Dublin, Ireland; [30]Dermatology Service, Hospital del Mar-IMAS, Barcelona, Spain; [31]Department of Dermatology and Venereology, Sahlgrenska University Hospital, Gothenburg, Sweden